

PARTITION-BASED IMAGE REPRESENTATION AS BASIS FOR USER-ASSISTED SEGMENTATION¹

F. Marqués
Univ. Politècnica Catalunya
08034 Barcelona
SPAIN

B. Marcotegui, F. Zanoguera
Ecole des Mines – CMM
77305 Fontainebleau
France

P. Correia
IT – IST
1049 – 001 Lisboa
Portugal

R. Mech, M. Wollborn
Univ. Hannover
30167 Hannover
Germany

ABSTRACT

This paper discusses the usefulness of a partition-based image representation in the context of user-assisted segmentation, and compares it with the two other most common image representations in this framework: pixel-based and transition-based. Partition-based image representations allow user-interaction at the region level, which is a very natural and friendly manner to interact with images. The paper describes various region-based implementations of the most common types of user interaction; namely, initial object selection, object refinement, imposing objects characteristics, selection of specific objects, selection of objects that are similar to a previously selected one. In all cases, partition-based image representations, and their associated region-based tools for user interaction, show up to be very suitable for interactive image segmentation.

1. INTRODUCTION

Nowadays, there is a growing interest on the creation of segmented contents from generic video sequences. Such an interest is mainly due to the emerging MPEG-4 and MPEG-7 standards as well as to the increasing demand of special effects in post-production of video data [3] [12].

Automatic video segmentation has shown to be a very cumbersome task, especially when no *a priori* information about the object(s) to be segmented is available. Thus, new segmentation techniques allowing user interaction are currently under study [1] [2] [5] [6] [7] [11].

Interaction has to be as intuitive and flexible as possible to enable the user helping and conducting the analysis process with a minimum effort. The main drawback of many interactive segmentation systems is that, since they rely on techniques developed for automatic applications, user interaction is not tailored to the user needs. The types of interaction that are commonly necessary cover the following aspects: initial object selection, object refinement when a new part of it becomes visible, imposing objects characteristics (e.g. its shape, type of motion, possible trajectory), selection of specific objects (e.g. human faces), selection of objects that are similar to a previously selected one.

In this paper we discuss the three main approaches of user interactive segmentation that have been proposed in the literature; namely, feature-based, contour-based and region-based. Each one of these approaches has an associated underlying image representation; pixel-based, transition-based and partition-based, respectively.

In the framework of partition-based image representation, we present techniques that allow addressing, in a natural and simple manner, the previous types of user interaction. Furthermore, we show that the most relevant features of the other two approaches can be adopted by the region-based interaction approach.

2. IMAGE REPRESENTATIONS AND INTERACTION APPROACHES

The three main types of interactive segmentation approaches arise from three different image representations. Here, the main attributes of each image representation and their associated interaction approaches are described. Commonly, the same principles that are applied for representing images are extended to sequence representation. Motion between consecutive frames is estimated and applied to the elements that form the representation (pixels, transitions or regions) to create a temporal link. Analogously, the interaction approaches defined for handling images are extended to the case of sequences, by allowing, for instance, trajectory specification.

Pixel-based image representation: It is the simplest image representation. In it, images are understood as a set of pixels with specific features but no connectivity aspects are taken into account in the image model. The associated user interaction approach is feature-based, where the user is asked to select a set of pixels in the image that correspond to the various texture- or motion-homogeneous parts of the objects to be segmented [2] [11]. Thus, the user is not required to precisely define the object, which is obtained by partitioning the feature space. Such a partition results in a classification of the image space.

Transition-based image representation: This type of representation divides the image into objects, which are defined by their boundaries. These boundaries correspond to transitions that mark the separation between neighboring homogenous areas. Since the relevant aspect is the location of such transitions, the related interaction approach is contour-based. In it, the user can roughly mark the position of the object boundaries and automatic algorithms accommodate them to the object real contours [5] [6]. This approach leads to smooth contours accurately representing the real object shape.

Partition-based image representation: In this type of representation, an image is understood as a collection of regions that totally covers it (a partition). Regions are homogeneous in the selected feature space and connected in the image space. Such an image representation enables region-

¹ This work was partially supported by the European Commission under the ACTS AC-098 MoMuSys project.

based user interaction. In it, the user can interact with the underlying partition(s) that represent the image.

3. REGION-BASED USER INTERACTION

Merging regions from an initial partition is the most straightforward region-based user interaction [1] [13]. However in this section we describe other much powerful techniques that cover the most common types of interaction.

3.1 Initial object selection

To obtain an accurate object definition, a fixed partition may not be good enough. Depending on the complexity of the object to be defined, the user may require a large number of mergings to create it. A hierarchy of partitions (nested partitions [13]) allowing a representation of the image at different levels of resolution is a very powerful starting point for assisted object definition. The different levels in the hierarchy can be associated to the number of regions contained in the partition. This way, the user can select the level of detail of the partition from which the merging of regions should start. In the technique proposed in [13], such a selection can be done in real time because a minimum spanning tree represents the nested partitions and all calculations are carried out on this synthetic data structure. Figure 1 shows a set of different partitions that can be selected as starting point for a given image.

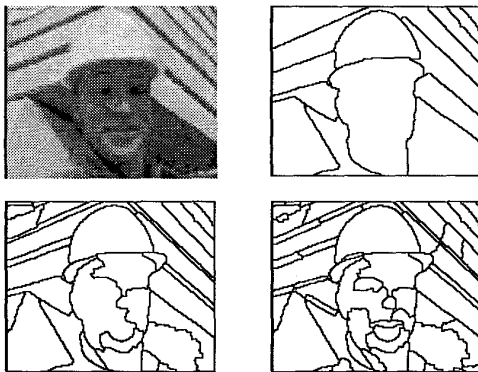


Figure 1 - Original image and three different levels of the hierarchy

To define the object, the user can also combine regions from different resolution levels, leading to region configurations that are not originally allowed in the minimum spanning tree. This is possible because the partitions are nested, that is, going up (down) in the hierarchy removes (adds) contours but does not create (remove) any new ones. In that way different resolution levels can be mixed in the same image, obtaining always a partition. Figure 2 shows an example of such type of interaction.

The common mechanisms for assisted object definition in feature-based or contour-based approaches can also be incorporated to the region-based approach. The set of selected pixels in a feature-based approach can be used as markers either to create the initial partition or to merge regions from a

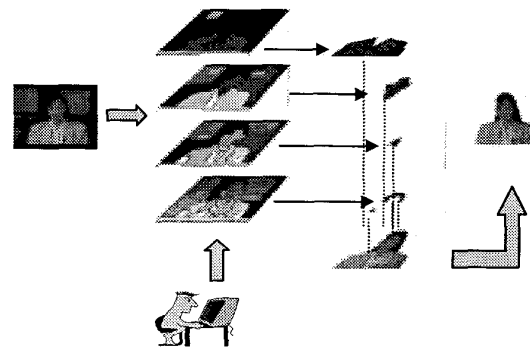


Figure 2 - Interactive segmentation based on a hierarchy of partitions

previously computed partition. This way, connectivity aspects can be combined with an object definition similar to the one used in feature-based approaches. For this purpose we can consider two different techniques. The set of pixels selected by a feature-based approach can be introduced as markers in the technique proposed in [13]. In this case, markers for the background as well as for each selected object are required. This set of pixels can also be introduced in a Binary Partition Tree discussed in [12]. In this case only markers for the selected objects are required, obtaining a segmentation of the image into separate objects that does not lead to a complete partition.

Analogously, the contour-based approach can also be extended to region-based by asking the user to roughly draw the interior and exterior boundaries of the selected object. These boundaries are used as markers in the segmentation process. Another possibility is to draw only one contour of a given width, that is a band around the real contour. The more the drawn contour is accurate, the thinner the band is. The inner and outer boundary of this band play the role of interior and exterior markers, and the user only has to give one contour. Such boundaries can again be used to guide the region merging up to obtaining the final object contour.

3.2 Object refinement

The previous idea of composing an object by adding regions from different resolution levels can be applied for refining a given object. Object refinement is especially useful in the case of tracking an initial object through successive frames. In this case, object refinement may be necessary if the object definition varies along the time or if the object tracking technique fails on providing a correct object shape.

The object tracking based on a partition-based approach tracks the selected object(s) as well as a finer partition of the whole image. Therefore, regions from the tracked partition can be added (removed) to the object. This is possible when using an object tracking technique based on a partition projection, such as that proposed in [8]. An object in a given image is represented by the partition of the image and a Look Up Table (LUT) that contains the information about which regions from the partition form the object. The previous partition information is projected into the current image and new regions are added, if

necessary, to create the current partition. The object shape in the current image is obtained by applying the information contained in the LUT. The algorithm has to make a decision on whether the new regions belong or not to the object. The set of new regions on which the decision is not straightforward can be proposed to the user as candidates for the interactive refinement.

The refinements described so far are assumed to be carried out on-line. However, an off-line refinement tool may also be necessary if, after processing a large sequence, a problem is detected and the user does not want to re-run the complete process. This possibility, which is not straightforward in feature-based or contour-based approaches, can be implemented in the region-based approach if the object shape is stored as the partition at each frame jointly with the LUT defining which regions from this partition belong to the object. This way, an off-line refinement process that allows adding (removing) a given region of the partition to (from) the object definition simply translates into an update of its associated LUT value in the successive partitions.

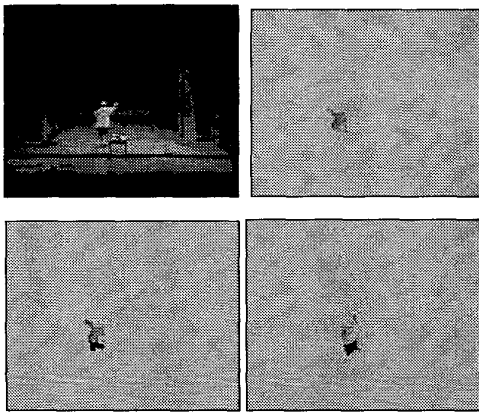


Figure 3 - Off-line re-definition of the object to be tracked.

Figure 3 presents an example of off-line interaction where the user, to compute the trajectory of the man in the scene, first selects to track his body. After processing the whole sequence the user decides to add the legs to the object to increase the robustness of the trajectory estimation. Instead of re-running the process, the user selects, from the stored partition, the region corresponding to the legs. The LUT is modified to assign this region to the object. Thanks to the partition projection the legs are added in one step to the whole sequence.

3.3 Specify object characteristics

As stated in section 2, contour-based approaches enable the introduction of constraints in the segmentation and tracking algorithms so that objects with smooth shapes are obtained. Region-based algorithms can also easily include *a priori* information in the analysis process and force the final object shape to have specific characteristics such as being smooth, formed by a given number of connected components or compact (without holes) [8]. Such an information can be introduced in

the creation of the projected partition or even in the final object definition. In this second case, notice that the improvement due to this information can be achieved on-line as well as off-line.

Information about the trajectory of the object or the motion in the scene can be introduced in the segmentation process, as well. The trajectory of the object to be segmented can be used in the partition projection step, allowing a more accurate tracking of the regions defining the object. In addition, if the object to be segmented is moving, motion information can be applied to perform an initialization or first estimation of its shape [10]. If the user knows that, in addition to a moving object, the sequence presents camera motion, the main motion can be estimated [10] and its information can be introduced the segmentation process.

3.4 Select specific objects

For common objects (typically, human faces, persons), the user should be able to semantically interact; that is, to directly ask for their detection, segmentation and tracking. Pixel- and transition-based image representations may allow obtaining rough approximations of faces based on, for instance, the color of the skin or the oval shape of a face, respectively. However, a partition-based image representation enables the natural combination of these attributes leading to a more robust analysis tool [9]. In Figure 4, an example of the detection of a face based on the use of partition-based image representations is presented.



Figure 4 - Example of face detection for sequence *Akiyo*.

3.5 Selection of similar objects

Once an object has been selected, several features (color, texture, motion...) can easily be extracted from it, following the same type of techniques as in feature-based approaches. The most homogeneous feature within the selected area will be used to detect similar objects in other images and introduced as markers in the region approach to find the real contours. Notice that in a region-based approach, feature estimation can be carried out in a more robust way since the set of sample points is enlarged due to the use of regions. In that way the algorithm learns from the user interaction and simplifies the detection of similar objects.

Another technique can be used to introduce the acquired knowledge about the object. In [4], the concept of detection of similar objects is applied to the tracking problem. Once the user has selected the object of interest, in that case, the helmet that has a very low contrast, the algorithm normalizes the following images so that the contrast in the boundary of the

object is maximized. This technique allows imposing local constraints to the segmentation algorithm and, therefore, to obtain partitions which are more suitable for object segmentation purposes.

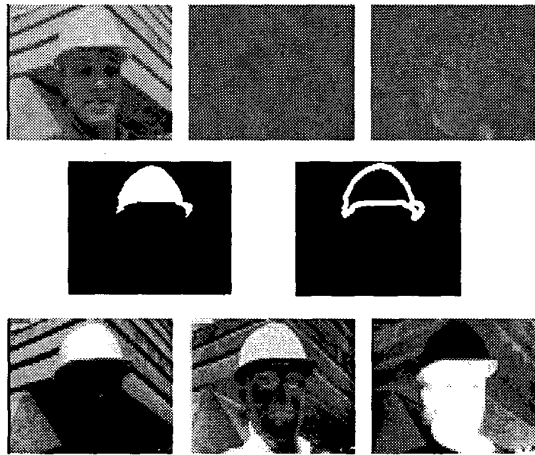


Figure 5 - Detection and tracking of similar objects.

The first row of Figure 5 presents the three-color components in the YUV color space. The second row shows the selected object and a band around its boundary. Finally, the third row shows the normalized color component images maximizing the contrast in the boundary of the selected object.

4. VOGUE: A REGION-BASED USER-ASSISTED SEGMENTATION SYSTEM

An example of user-assisted segmentation system that incorporates some of the previous region-based user interaction possibilities is the so-called VOGUE (Video Object Generator tool with User Environment) [7]. In it, three different region-based interactive algorithms have been combined in the same graphic user interface; namely, spatial segmentation [13], object tracking [8] and temporal segmentation [10]. VOGUE is a clear example of the feasibility and user-friendliness of a partition-based interactive segmentation approach.

5. CONCLUSION

In this paper we show that a partition-based image representation is very suitable for interactive image segmentation. It offers a flexible framework for the commonly required types of interaction: object definition and refinement, selection of a specific type of object, etc. Moreover the interaction is user friendly and does not require a deep knowledge of the algorithms by the user, i.e. no parameters to be tuned. Finally the partition approach can adopt the most relevant features of the other two approaches discussed in Section 2.

However, it has to be noticed that some interesting functionalities of contour-based approaches have not been adopted by partition-based approaches, yet. This is the case of the possibility of fixing a specific point in the contour of an object during the tracking process (for instance, a corner that must be preserved) or to easily enforce the object shape to have a predetermined form. Such aspects are still open for research.

REFERENCES

- [1] R. Castagno, T. Ebrahimi and M. Kunt, "Video segmentation based on multiple features for interactive multimedia applications", *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 8, num. 5, pages 562-571, September 1998.
- [2] E. Chalom and M. Bove, "Segmentation of an image sequence using multi-dimensional image attributes", *Proceedings of the ICIP 96*, pp. 525-528, Lausanne, Switzerland, 1996.
- [3] P. Correia, F. Pereira, "The Role of Analysis in Content-Based Video Coding and Indexing", *Signal Processing*, 66 (1998), 125-142.
- [4] C. Gomila and F. Meyer, "Automatic Video Object Generation Tool: segmentation and tracking of persons in real time". In the *Annals of Telecommunications* May 2000.
- [5] C. Gu, M. C. Lee, "Semiautomatic segmentation and tracking of semantic video objects", *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 8, num. 5, pages 572--584, September 1998.
- [6] S. Kruse, X. Bardella, F. Schweitzer, M. Valero, "An interactive image segmentation scheme", *Picture Coding Symposium*, pp. 169-173, Portland, USA, 1998.
- [7] B. Marcotegui et al., "A video object generation tool allowing friendly user interaction", *Proceedings of the ICIP 99*, paper 26AP4, Kobe, 1999.
- [8] F. Marqués, J. Llach, "Tracking of generic objects for Video Object generation", *Proceedings of the ICIP 98*, pp. 628-632, Chicago, 1998.
- [9] F. Marqués, V. Vilaplana, A. Buxes, "Human face segmentation and tracking using connected components and partition projection", *Proceedings of the ICIP 99*, paper 27PO303, Kobe, Japan, 1999.
- [10] R. Mech and M. Wollborn, "A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera", *Signal Processing*, Vol. 66, No. 2, pp. 203-217, April 1998.
- [11] N.E. O'Connor and S. Marlow, "Supervised Semantic Object Segmentation and Tracking via EM-based Estimation of Mixture Density Parameters". *Noblesse Workshop on Non-Linear Model Based Image Analysis*. Glasgow, July 1998.
- [12] P. Salembier, F. Marqués, "Region-based representations of image and video: Segmentation tools for multimedia services", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, n. 8, pp. 1147-1169, Dec. 1999.
- [13] F. Zanoguera, B. Marcotegui and F. Meyer, "A toolbox for interactive segmentation based on nested partitions". In *Proceedings of the ICIP'99*, Kobe, Japan, Oct 99.